The EMBO Journal vol.15 no.18 pp.5022–5031, 1996

hTAF$_{II}$68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II

Anne Bertolotti, Yves Lutz, David J.Heard, Pierre Chambon and Laszlo Tora[1]

Institut de Génétique et de Biologie Moléculaire et Cellulaire CNRS/INSERM/ULP, Collège de France BP 163, F-67404 Illkirch Cedex, CU de Strasbourg, France

[1]Corresponding author

TFIID is the main sequence-specific DNA-binding component of the RNA polymerase II (Pol II) transcriptional machinery. It is a multiprotein complex composed of the TATA-binding protein (TBP) and TBP-associated factors (TAF$_{II}$s). Here we report the cloning and characterization of a novel human TBP-associated factor, hTAF$_{II}$68. It contains a consensus RNA-binding domain (RNP-CS) and binds not only RNA, but also single stranded (ss) DNA. hTAF$_{II}$68 shares extensive sequence similarity with TLS/FUS and EWS, two human nuclear RNA-binding pro-oncoproteins which are products of genes commonly translocated in human sarcomas. Like hTAF$_{II}$68, TLS/FUS is also associated with a sub-population of TFIID complexes chromatographically separable from those containing hTAF$_{II}$68. Therefore, these RNA and/or ssDNA-binding proteins may play specific roles during transcription initiation at distinct promoters. Moreover, we demonstrate that hTAF$_{II}$68 co-purifies also with the human RNA polymerase II and can enter the preinitiation complex together with Pol II.

Keywords: ssDNA-binding protein/RNA polymerase II/specific TAF$_{II}$s/transcription initiation/translocation

Introduction

Transcription can be divided into several processes that include promoter recognition and binding, preinitiation complex (PIC) assembly, RNA chain initiation, RNA chain elongation and termination. Initiation of transcription by RNA polymerase II (Pol II) from a core promoter region requires the coordinated action of several initiation factors which form a PIC (Serizawa *et al.*, 1994). Assembly of the PIC is nucleated by the sequence-specific binding of the TATA-binding protein (TBP) to the core promoter. TBP is the main DNA-binding subunit of the TFIID multiprotein complex. TFIIB then interacts with TFIID (or TBP)–DNA complex allowing the subsequent entry of TFIIF and Pol II to form the DBPolF complex (Zawel *et al.*, 1995). A systematic *in vitro* study of 14 supercoiled core promoters indicated that with the exception of the immunoglobulin heavy chain promoter, which required only TBP, TFIIB and Pol II, all other promoters tested required, in addition, TFIIF for basal transcription (Parvin *et al.*, 1994). Transcription from promoter templates with relaxed DNA topology is further dependent on TFIIE and TFIIH (Timmers, 1994; Holstege *et al.*, 1995). In contrast to the sequential assembly model of PIC formation the recent discovery of the Pol II holoenzyme suggests that the initiation complex may in fact exist in the cells as a superassembly of Pol II, initiation factors, mediators and coactivators, all physically linked through a network of protein–protein interactions (Kim *et al.*, 1994; Hengartner *et al.*, 1995; Ossipow *et al.*, 1995).

The exact molecular mechanism by which upstream activators stimulate transcription is not known. TBP alone is sufficient for basal level transcription *in vitro*, whereas the multisubunit TFIID complex(es) containing TBP and up to 13 TBP-associated factors (TAF$_{II}$s) (Pugh *et al.*, 1991; Tanese *et al.*, 1991; Timmers *et al.*, 1992; Zhou *et al.*, 1992; Brou *et al.*, 1993a; Chiang *et al.*, 1993; Yokomori *et al.*, 1993; Jacq *et al.*, 1994; Mengus *et al.*, 1995) is required for activated transcription. Transcriptional activation experiments, using either partially assembled TFIID complexes or antibodies raised against TAF$_{II}$s, confirmed that TAF$_{II}$s are required for activated transcription and have demonstrated that different transcriptional activation domains interact with distinct TAF$_{II}$s (Goodrich *et al.*, 1993; Hoey *et al.*, 1993; Chen *et al.*, 1994; Jacq *et al.*, 1994; Sauer *et al.*, 1995). In addition, TAF$_{II}$s have also been shown to be involved in promoter recognition (Verrijzer *et al.*, 1994) and in protein–protein interactions with general transcription factors (Goodrich *et al.*, 1993; Histake *et al.*, 1995; Klemm *et al.*, 1995; Dubrovskaya *et al.*, 1996).

Previously, we have shown that functionally distinct TFIID complexes composed of both common and specific TAF$_{II}$s exist in human HeLa cells (Brou *et al.*, 1993a; Jacq *et al.*, 1994; Mengus *et al.*, 1995). Core TAF$_{II}$s are present in each TFIID complex, whereas specific TAF$_{II}$s can only be found in TFIID sub-populations often in substoichiometric amounts as compared with TBP (Brou *et al.*, 1993a; Jacq *et al.*, 1994). By immunoprecipitating TFIID complexes from chromatographically separable fractions, which eluted from a phosphocellulose column (PC) at different salt concentrations (0.3, 0.5 and 1.0 M KCl), we were able to show that the different TFIID complexes exhibited functionally distinct properties. Factors associated with the PC1.0-derived TFIID sub-population were required for mediating transcriptional stimulation by the activation functions of the oestrogen receptor (ER), VP16 and transcriptional enhancer factor 1 (TEF-1), while the PC0.3-derived TFIID sub-population mediated only TEF-1 activation (Brou *et al.*, 1993a). For example, a specific TAF$_{II}$, hTAF$_{II}$30, was identified from HeLa cells and shown to be present in a subset of TFIID complexes (Brou *et al.*, 1993a; Jacq *et al.*, 1994). hTAF$_{II}$30 interacts specifically with the hormone-binding domain of the ER (Jacq *et al.*, 1994). In support of our findings, several specific TAF_{II}s have also been described in a distinct *Drosophila* (d) TFIID complex (Hansen *et al.*, 1995). These specific $dTAF_{II}$s appear to be substoichiometric when compared with the identified $dTAF_{II}$s or TBP and are thought to direct the promoter selectivity of TFIID during transcription (Hansen *et al.*, 1995). Together the studies described above highlight the possibility that distinct TFIID complexes may have specific roles in recognition of different promoters and/or interaction with transcriptional activators dependent on their TAF_{II} composition.

Transcription is also controlled subsequent to initiation. Different processes can be regulated during the elongation reaction, such as transcriptional pausing, transcript cleavage and reactivation, and post-translational modifications of Pol II (for review see Kane, 1994), resulting in an increase in the processivity of the polymerase (Yankulov *et al.*, 1994). These observations imply that in the presence or absence of activators different types of initiation events occur resulting in the assembly of processive or non-processive transcription elongation complexes, respectively

Here we report the cloning and characterization of a novel specific TAF_{II}, $hTAF_{II}68$, that was purified on the basis of its association with a distinct TFIID subpopulation. $hTAF_{II}68$ shares extensive sequence similarity with TLS/FUS and EWS, two human nuclear RNA-binding pro-oncoproteins which are products of genes commonly translocated in human sarcomas (Delattre *et al.*, 1992; Crozat *et al.*, 1993). The consensus RNA-binding domain of $hTAF_{II}68$ and its surrounding sequences bind not only RNA but also single stranded (ss) DNA. Antibodies raised against $hTAF_{II}68$ co-immunoprecipitate TFIID and vice versa, anti-TBP or anti-$TAF_{II}100$ mAbs co-immunopurify $hTAF_{II}68$. The structural homology between $hTAF_{II}68$ and TLS/FUS may also be functional, as TLS/FUS is also associated with TFIID complexes which are distinct from those containing $hTAF_{II}68$. Moreover, $hTAF_{II}68$ is also associated with the human RNA Pol II and is able to enter into the PIC together with Pol II. Our results suggests that $hTAF_{II}68$ could play a role during both transcription initiation and elongation.

Results

Cloning and expression of a human TFIID-associated protein

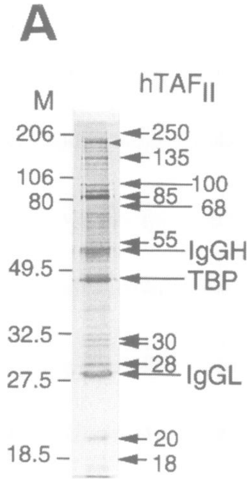
To understand better the function of the distinct TFIID sub-populations we have further characterized novel proteins from the same TFIID preparation that allowed us to identify and clone the specific TAF_{II}, $hTAF_{II}30$ (Jacq *et al.*, 1994). A 68 kDa polypeptide (hereafter called $hTAF_{II}68$) which is stably associated with this TFIID complex in high stringency salt washes (700 mM KCl), although in substoichiometric amounts when compared with TBP or other core TAF_{II}s (Figure 1A), was subjected to microsequencing. The amino acid sequences of three tryptic peptides (Figure 1B) were used to design degenerate oligonucleotides for screening a HeLa cell cDNA library. A cDNA clone was isolated containing an open reading frame (ORF) encoding a protein of 589 amino acids with a predicted molecular mass of 65 kDa (Figure 1B). The three peptide sequences (Figure 1B) derived from the tryptic digests of the endogenous protein were found within this ORF. To determine whether the isolated cDNA encodes a protein with an apparent molecular mass identical to that of endogenous $hTAF_{II}68$, *in vitro* transcription–translation was performed and the protein product tested by Western blot analysis using an anti-$TAF_{II}68$ mouse antiserum. *In vitro* translated $hTAF_{II}68$ had the same mobility in SDS–PAGE as the endogenous HeLa cell $hTAF_{II}68$ (Figure 3A, lane 1 and 2). These results together confirm the identity of the isolated cDNA.

An amino acid similarity search using the Pearson and Lipman method (Pearson *et al.*, 1988) revealed that $hTAF_{II}68$ shares extensive sequence similarity with TLS/FUS (74%) and EWS (70%), two human nuclear RNA-binding proteins which are products of genes often translocated in human sarcomas (Delattre *et al.*, 1992; Crozat *et al.*, 1993). Similar to TLS/FUS and EWS, the amino acid sequence of $hTAF_{II}68$ can be divided into three putative functional domains. The N-terminal 160 amino acid region is rich in glutamine (21%), serine (19%) and tyrosine (15%). The corresponding regions of TLS/FUS and EWS have been shown to function as transcriptional activating domains (Ohno *et al.*, 1993, 1994; Prasad *et al.*, 1994; Sanchez-Garcia *et al.*, 1994). The central region of $hTAF_{II}68$ contains a conserved 87 amino acid RNA-binding domain, called the RNP-motif or RNP-CS (boxed in Figure 1B; and see Kenan *et al.*, 1991; Burd *et al.*, 1994). When compared with the family of RNP-motif-containing proteins (Figure 1C), the $hTAF_{II}68$ RNP shares the highest degree of similarity with those of human TLS/FUS (84% identity) and EWS (59% identity), and with that of the *Drosophila* Cabeza protein (52% identity; Stolow *et al.*, 1995). The C-terminal 270 amino acid region of $hTAF_{II}68$ contains 20 Arg-Gly-Gly (RGG) repeats, initially described as an RNA-binding motif in the hnRNP-U protein (Kiledjian *et al.*, 1992), and typically found in combination with other types of RNA-binding domains (Burd *et al.*, 1994). However, in $hTAF_{II}68$ the RGG repeats are not as closely spaced as in other RNA-binding proteins (Kiledjian *et al.*, 1992). This C-terminal region also contains 17 perfect and three imperfect closely spaced Gly-Gly-Tyr-Gly-Gly-Asp-Arg repeats (GGYGGDR; bold characters underlined in Figure 1B). These GGYGGDR repeats have not yet been described but, since they overlap the RGG repeats, may define a new RNA-binding motif.

$hTAF_{II}68$ binds RNA and ssDNA

Next we determined whether the putative RNA-binding motif(s) was able to confer RNA-binding activity to $hTAF_{II}68$. To test this, the full-length $hTAF_{II}68$ (tagged with six histidines) was overexpressed in SF9 cells using the baculovirus expression system (see Materials and methods), and a truncated version of $hTAF_{II}68$, overlapping the RNP-motif, was overexpressed in *Escherichia coli* as a glutathione *S*-transferase (GST) fusion protein. The purified proteins (see Figure 2A, lanes 1–3) were then used in North- and South-Western blots with RNA, ssDNA and dsDNA probes. Labelled RNA bound both to the recombinant His-$TAF_{II}68$ and the GST–RNP fusion protein, but not to the GST control (Figure 2A, lanes 4–6). Interestingly, the ssDNA probe, containing the AdMLP sequence from –40 to –11, also bound efficiently to both the full-length $hTAF_{II}68$ and the RNP-motif-containing

A



B

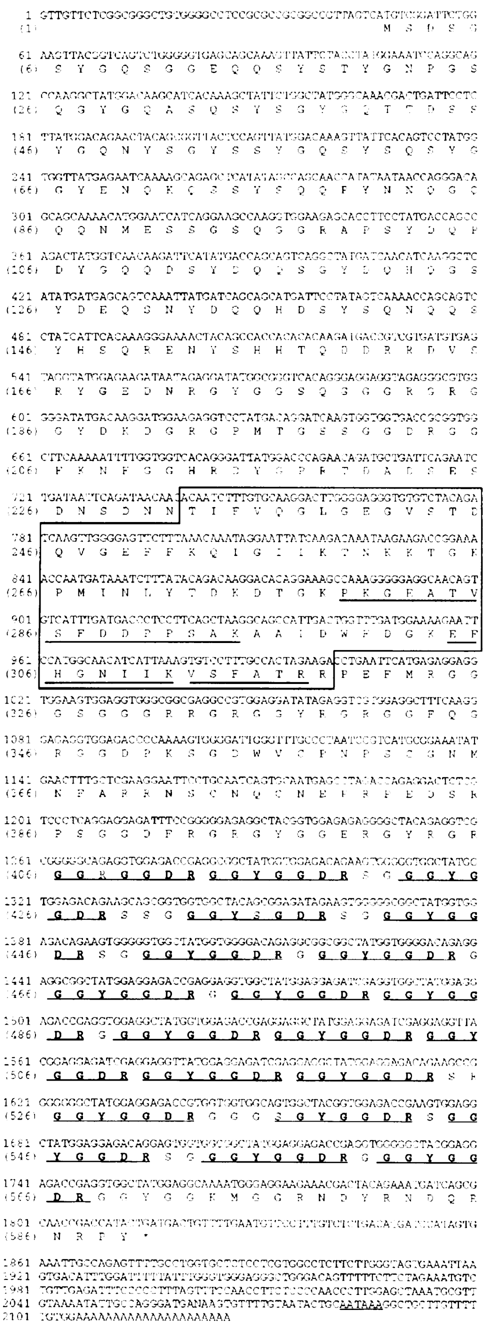
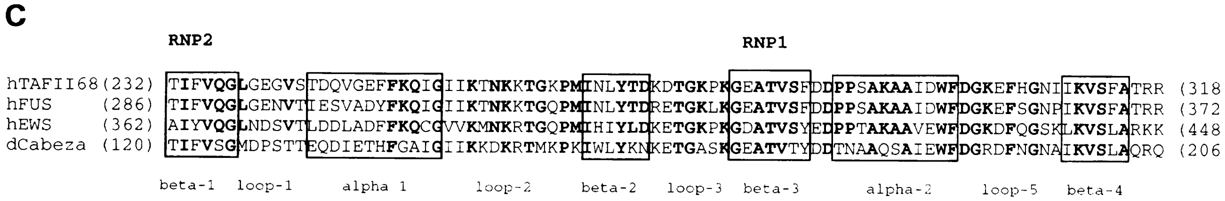


Fig. 1. Identification of hTAF$_{II}$68. (**A**) TFIID complex from the phosphocellulose 1.0 M KCl fraction (PC1.0 TFIID) was immunoprecipitated with the anti-TBP mAb 3G3 (Brou *et al.*, 1993a), the protein G–mAb bound TFIID complexes were washed with 700 mM KCl containing IP buffer, eluted with the corresponding epitope peptide (Brou *et al.*, 1993a), resolved by SDS–PAGE and silver stained. TBP and known human Pol II TAFs (hTAF$_{II}$s) including hTAF$_{II}$68 are indicated (Ruppert *et al.*, 1993; Weinzierl *et al.*, 1993; Jacq *et al.*, 1994; Chiang *et al.*, 1995; Mengus *et al.*, 1995). A contaminating protein species that co-purified with the mAb is labelled with an arrowhead. (**B**) Nucleotide sequence of the human TAF$_{II}$68 cDNA clone and the corresponding amino acid sequence is shown. The hTAF$_{II}$68 ORF extends from nucleotide 47 to 1814. The sequence element corresponding to a potential polyadenylation signal is underlined. The amino acid sequences determined by microsequencing of peptides obtained after tryptic digestion of the endogenous hTAF$_{II}$68 protein are underlined. The RNP-motif (Burd *et al.*, 1994) of hTAF$_{II}$68 is boxed. The Gly-Gly-Tyr-Gly-Gly-Asp-Arg (GGYGGDR) repeats are shown with underlined bold characters. (**C**) Alignment of the amino acid sequences of human (h) TAF$_{II}$68, TLS/FUS (Crozat *et al.*, 1993), EWS (Delattre *et al.*, 1992) and *Drosophila* (d) Cabeza protein (Stolow *et al.*, 1995) in the region corresponding to the conserved RNA-binding domains or RNP-motifs (Kenan *et al.*, 1991; Burd *et al.*, 1994). Identical amino acids in the three human proteins are shown with bold characters. Secondary structural elements are indicated according to Nagai *et al.* (1990), the four β sheets and the two α helices are boxed.

C



mutant (lanes 7 and 8), whereas the corresponding dsDNA probe did not (lanes 10–12). Thus, the RNA-binding motif of hTAF$_{II}$68 and its surrounding sequences has both RNA and ssDNA-binding specificity.

The specificity of the ssDNA-binding activity of hTAF$_{II}$68 was further analysed by electrophoretic mobility shift assay (EMSA). The GST–RNP fusion protein forms complexes with the same labelled ssDNA probe that was used above in the South-Western blot experiment (Figure 2B, lane 2). These protein–ssDNA complexes are super-shifted by the addition of anti-hTAF$_{II}$68 mAb (lane 3) indicating the specificity of these complexes. The GST–

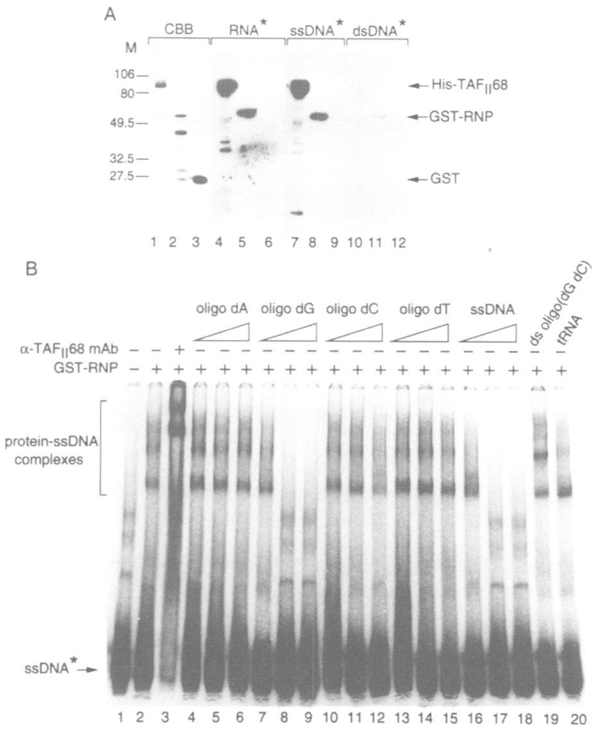


Fig. 2. hTAF$_{II}$68 binds RNA and ssDNA. (**A**) The full-length hTAF$_{II}$68 was overexpressed in the baculovirus system (His-TAF$_{II}$68) and its RNP-motif as a GST fusion protein (GST–RNP), together with the GST control, in *E.coli*. The purified proteins (1 μg of each) were resolved by SDS–PAGE and stained by Coomassie Blue (CBB, lanes 1–3). The same samples were tested by North- and South-Western blotting using a labelled RNA probe containing the rabbit β-globin gene sequences from –9 to +290 (lanes 4–6), labelled ssDNA probe containing the coding strand of the Adenovirus Major Late promoter (AdMLP) from –40 to +13 relative to the start site (lanes 7–9) and labelled dsDNA probe containing both strands of the AdMLP from –40 to +13 (lanes 10–12). (**B**) Analysis of DNA-binding properties of hTAF$_{II}$68 by EMSA. The GST–RNP fusion protein was incubated with a radiolabelled ssDNA probe, containing the coding strand of the Adenovirus Major Late promoter (see panel A) to form protein–ssDNA complexes (indicated). Anti-TAF$_{II}$68 mAb 2B10 was used to supershift and confirmed the specificity of these complexes. The binding preference of the protein was tested by preincubating GST–RNP fusion protein with increasing amounts of unlabelled deoxyoligonucleotide homopolymers (24 nt) as indicated (oligo-dA, -dG, -dC, -dT: lanes 4–15), with unlabelled ssDNA probe (lanes 16–18) or with the highest amount of ds oligo(dG/dC) (lane 19) or tRNA (lane 20). The pre-formed complexes were then challenged by the addition of labelled ssDNA probe and subjected to PAGE.

RNP fusion protein was then preincubated with unlabelled deoxyoligonucleotide homopolymers prior to the addition of the ssDNA probe. Increasing amounts of oligo(dG) significantly inhibited the binding of the GST–RNP fusion protein to the probe (lanes 7–9), almost as efficiently as the unlabelled ssDNA probe itself (lanes 16–18). Oligo(dC) also has a very weak effect on binding of the GST–RNP fusion protein to the ssDNA probe, but only at the highest molar excess (2000-fold, lane 12). In contrast, no inhibition of binding was observed when the GST–RNP fusion protein was preincubated with either oligo(dA) or oligo(dT) (lanes 4–6 and 13–15). In addition, 2000 molar excess of either double stranded oligo(dGdC) or tRNA has only a weak effect on the binding of the protein to the ssDNA probe (lanes 19 and 20). In the presence of ds oligo(dGdC) and tRNA the non-specific background is less apparent suggesting that these molecules behave as non-specific competitors in this experiment (note, that in the binding studies shown in Figure 2A 20 μg/ml tRNA was included as non-specific competitor). The ssDNA probe used in the gel shift experiment corresponds to the AdMLP sequence from –40 to –11 which contains 38% of G. The GST–RNP fusion protein forms at least three complexes with the single stranded probe suggesting that more than one molecule of the RNP-containing protein can bind to the probe. This is in good agreement with the fact that the probe contains three dG stretches and that the hTAF$_{II}$68 RNP-motif has the strongest affinity to oligo(dG) sequences among the different homopolymeric DNAs tested.

hTAF$_{II}$68 and TLS/FUS are present in distinct TFIID complexes

Since hTAF$_{II}$68 is substoichiometrically present in the PC1.0-derived TFIID complex, the original PC chromatographic fractions were tested for the presence of hTAF$_{II}$68 by Western blot analysis using the anti-TAF$_{II}$68 mouse monoclonal antibody (mAb2B11). Interestingly, most of the hTAF$_{II}$68 segregates away from the PC1.0 fraction (Figure 3B, lanes 1–3). To verify the specific association of hTAF$_{II}$68 with the PC1.0-derived TFIID complex, distinct TFIID complexes were immunopurified with an anti-TBP mAb from the three different PC fractions (Brou *et al.*, 1993a; Jacq *et al.*, 1994). The eluted TFIID complexes (Brou *et al.*, 1993a) were tested using the anti-TAF$_{II}$68 mAb (2B11). In agreement with the original purification data, most of the hTAF$_{II}$68 was present in the PC1.0-derived TFIID complex (Figure 3B, lane 4). All the PC fractions contain TBP and hTAF$_{II}$68 (see Figure 3B, lanes 1–3 and 3C) however hTAF$_{II}$68 co-immunoprecipitates with TBP only from the PC1.0 fraction indicating that hTAF$_{II}$68 is specifically associated with the PC1.0-derived TFIID. Note, that the three different TFIID complexes contain approximately the same amount of hTBP and hTAF$_{II}$100 (Figure 3C), and function in Pol II transcription (Jacq *et al.*, 1994). To further confirm the association of hTAF$_{II}$68 with the TFIID complex(es), the PC1.0 fraction was immunoprecipitated with the anti-TAF$_{II}$68 mAb (2B10). This antibody recognized the native endogenous hTAF$_{II}$68, and it co-immunoprecipitated components of the TFIID complexes such as hTBP and hTAF$_{II}$100 (Figure 3D, lane 1; and data not shown). These results confirm that hTAF$_{II}$68 is tightly associated with TFIID.

The sequence similarity shared by hTAF$_{II}$68, TLS/FUS and EWS suggests that these proteins may all be present in TFIID complexes. To test this possibility an anti-TLS/FUS mAb that specifically recognized TLS/FUS in a HeLa cell nuclear extract (Figure 3B, lane 10) was raised and used to investigate whether any of the immunopurified TFIID complexes described above (Figure 3C; and see also Brou *et al.*, 1993a; Jacq *et al.*, 1994) contain TLS/FUS. Western blot analysis demonstrated that TLS/FUS was associated with TFIID complex (Figure 3B, lane 9). Interestingly, TLS/FUS was specifically detected in the PC0.3-derived TFIID complex but was absent from both the PC0.5- and the PC1.0-derived TFIID complexes (Figure 3B, compare lanes 7–9). To confirm that TLS/FUS and hTAF$_{II}$68 are truly associated with TFIID com-

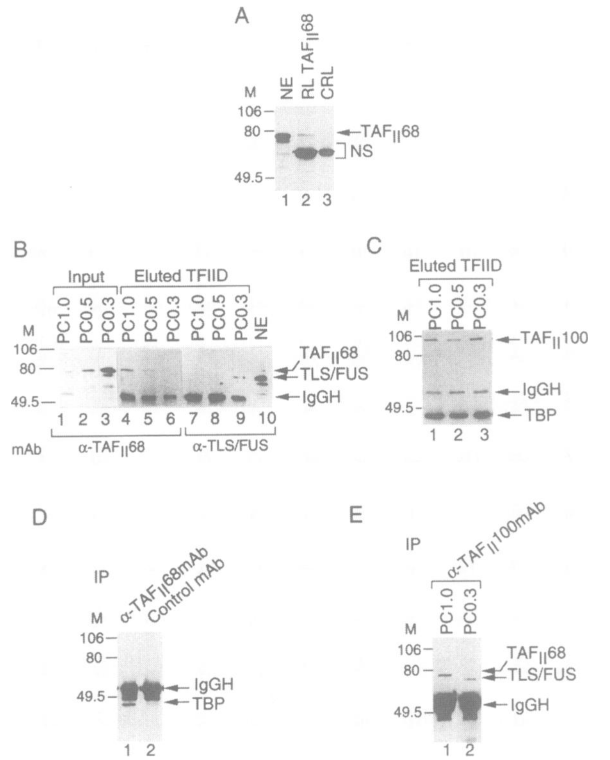


Fig. 3. hTAF$_{II}$68 and TLS/FUS are present in distinct TFIID complexes. (**A**) The hTAF$_{II}$68 cDNA was *in vitro* transcribed/translated in a rabbit reticulocyte lysate system (lane 2) and compared with the endogenous hTAF$_{II}$68 from a HeLa cell nuclear extract (NE, lane 1) by Western blot analysis using a mouse antiserum raised against a 17 amino acid peptide (amino acids 213 to 229, see Figure 1B). Note that the antiserum revealed a non-specific protein species (NS) in both the TAF$_{II}$68 containing rabbit reticulocyte lysate (lane 2) and in the control reticulocyte lysate (CRL, lane 3). (**B**) Fractions eluting from the PC column at 0.3 (PC0.3), 0.5 (PC0.5) and 1 M (PC1.0) KCl concentrations (see also Figure 5A) were analysed by Western blot using an anti-hTAF$_{II}$68 mAb (lanes 1–3). TFIID complexes were immunopurified from the PC1.0 (lane 1), PC0.5 (lane 2) and PC0.3 (lane 3) fractions (Input) with an anti-TBP mAb 3G3. The TFIID complexes were then eluted with an excess of the corresponding epitope peptide (Brou *et al.*, 1993a) and analysed by Western blot using an anti-hTAF$_{II}$68 mAb (lanes 4–6) or an anti-TLS/FUS mAb (lanes 7–9). The anti-TLS/FUS mAb recognizes the endogenous TLS/FUS protein in HeLa cell nuclear extract (NE; lane 10) and does not cross-react with hTAF$_{II}$68 (compare lanes 4 and 7). (**C**) The three different TFIID complexes analysed in panel B were tested by Western blot using the anti-TBP mAb3G3 and the anti-hTAF$_{II}$100 mAb 2D2. (**D**) The anti-hTAF$_{II}$68 mAb immunoprecipitates TBP-containing complexes. Phosphocellulose 1.0 M KCl fraction was immunoprecipitated by using either an anti-hTAF$_{II}$68 mAb (lane 1) or an unrelated control mAb (lane 2). After five washes with 700 mM KCl containing IP buffer, resin-bound proteins were analysed by Western blot with an anti-TBP mAb 3G3. (**E**) The anti-hTAF$_{II}$100 mAb 2D2 immunoprecipitates both hTAF$_{II}$68 and TLS/FUS. Phosphocellulose 1.0 M KCl (lane 1) and 0.3 M KCl (lane 2) fractions were immunoprecipitated by using the anti-hTAF$_{II}$100 mAb 2D2, resin-bound proteins were resolved by SDS–PAGE and analysed by Western blot with an anti-TAF$_{II}$68 mAb and an anti-TLS/FUS mAb. In (A–C) and (E) the immunoglobulin heavy chain is indicated (IgGH).

plexes, and not with free TBP or other TBP-containing complexes, immunoprecipitation was performed on the PC0.3 and the PC1.0 fractions using an antibody against hTAF$_{II}$100, a core TAF$_{II}$ present in all hTFIID complexes characterized to date (Dubrovskaya *et al.*, 1996). Both hTAF$_{II}$68 and TLS/FUS are immunoprecipitated by the anti-hTAF$_{II}$100 mAb 2D2 (Figure 3E, lane 1 and 2). Thus, TLS/FUS and hTAF$_{II}$68 are indeed associated with distinct TFIID complexes and may therefore contribute to the functional differences observed for these complexes (Brou *et al.*, 1993a; Jacq *et al.*, 1994). It remains to be tested whether EWS, the third member of the TET (TLS/FUS, EWS, TAF$_{II}$68) family, is also associated with TFIID.

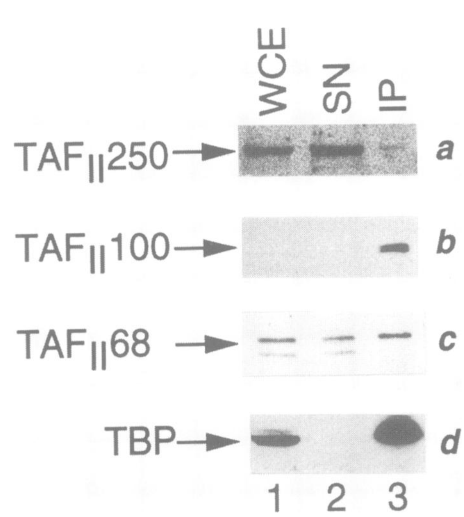


Fig. 4. Comparison of the extent of association of different TAF$_{II}$s with TBP. Proteins from a HeLa whole cell extract (WCE) were immunoprecipitated with the anti-TBP mAb 2C1. The WCE (lane 1), the supernatant of the immunoprecipitation (SN; lane 2) and the TBP-containing complexes (IP; lane 3) were separated by SDS–PAGE, blotted and the same blot tested successively with antibodies raised against (**a**) hTAF$_{II}$250 (Sekiguchi *et al.*, 1991), (**b**) hTAF$_{II}$100 (1TA), (**c**) hTAF$_{II}$68 and (**d**) TBP (3G3).

hTAF$_{II}$68 is also associated with the human RNA Pol II complex

Since hTAF$_{II}$68 appeared not to be exclusively associated with TFIID (see Figure 3B, lanes 1–6), the extent of association of different TAF$_{II}$s with TFIID complexes was compared with that of hTAF$_{II}$68. TBP-containing complexes were immunoprecipitated from HeLa whole cell extracts (WCE) with the anti-TBP mAb (3G3). The WCE, the supernatant of the immunoprecipitation and the TBP-containing complexes were analysed by Western blot with anti-TBP and different anti-TAF$_{II}$ antibodies. As shown in Figure 4, all of hTAF$_{II}$100 and TBP are immunoprecipitated by mAb 3G3 (panels b and d). In contrast, 90–95% of the total hTAF$_{II}$250 (CCG1) was found in TBP-free fractions (Figure 4, panel a). Similarly, only 5–10% of the cellular hTAF$_{II}$68 (and TLS/FUS; data not shown) is associated with TBP-containing complexes (Figure 4, panel c). These results confirm that hTAF$_{II}$68 is not exclusively associated with TFIID and suggest that hTAF$_{II}$68 may have several roles in the cell.

Thus, a number of chromatographic fractions obtained during the purification of different Pol II transcription factors were tested for the presence of hTAF$_{II}$68 (Figure 5A and data not shown). This analysis showed that the significant portion of hTAF$_{II}$68 that is present in the PC0.5 fraction (see Figure 3B, lane 2 and 5B, lane 3) co-purifies with Pol II that is both free from other basal factors and active in a reconstituted transcription system (Dubrovskaya *et al.*, 1996). hTAF$_{II}$68 co-purifies with the RNA polymerase II over five columns as determined by Western

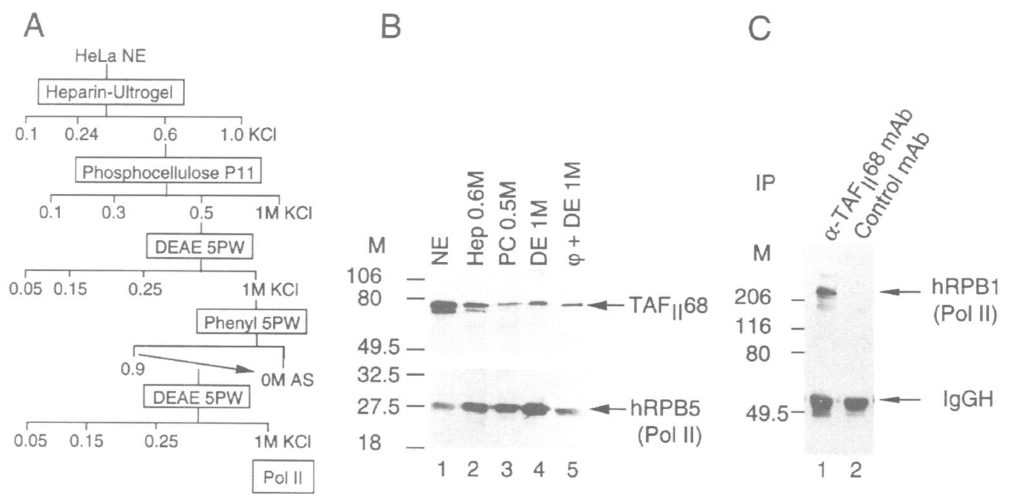


Fig. 5. (**A**) The chromatography protocol used to purify Pol II is outlined. The concentrations of KCl and ammonium sulphate (AS) used in column elutions are indicated. (**B**) hTAF$_{II}$68 co-purifies with Pol II. The different chromatographic fractions (see Materials and methods and panel A) resulting in approximately the same Pol II activity, in a reconstituted transcription system, were tested by Western blot with an anti-TAF$_{II}$68 mAb and a mAb raised against the 25 kDa subunit of Pol II (hRPB5). Hep, heparin-Ultrogel column; PC, phosphocellulose P11 column; DE, DEAE 5PW HPLC column; ϕ, phenyl-5PW HPLC column. (**C**) Anti-TAF$_{II}$68 mAb immunoprecipitates the Pol II complex. 500 μl of HeLa cell nuclear extract was immunoprecipitated with 20 μg of mAb2B10 or unrelated mAb (2GV3; White *et al.*, 1992) bound to 100 μl of protein G resin. After five washes with IP buffer containing 1 M KCl, 5 μl of the resin was boiled and proteins were resolved by SDS–PAGE and analysed by Western blot with an anti-CTD antibody. The position of the largest subunit of Pol II (hRPB1) and the immunoglobulin heavy chain is indicated (IgGH).

blot using an antibody raised against the 25 kDa subunit of Pol II (hRPB5) and an anti-TAF$_{II}$68 mAb (Figure 5B). To verify further the association of hTAF$_{II}$68 with Pol II, hTAF$_{II}$68 was immunopurified from HeLa cell nuclear extracts with an anti-TAF$_{II}$68 mAb. The complexes were washed at 1 M KCl and hTAF$_{II}$68-bound proteins were analysed by Western blot (Figure 5C). The anti-TAF$_{II}$68 mAb clearly co-immunoprecipitated Pol II (Figure 5C, lane 1), as detected using an antibody raised against the C-terminal domain (CTD) of the largest subunit of Pol II (hRPB1), while the control unrelated mAb did not (lane 2). This interaction was resistant to nuclease treatments indicating that this is a direct protein–protein interaction with the Pol II complex (data not shown). These results show that hTAF$_{II}$68 is tightly associated with the Pol II.

hTAF$_{II}$68 incorporates into the PIC together with Pol II

To determine whether hTAF$_{II}$68 is present in the PIC, or enters only into the ternary transcription elongation complex, EMSAs were carried out with ^{32}P-labelled oligonucleotides encompassing the AdMLP TATA box. Using recombinant (TBP, TFIIA and TFIIB) or purified (TFIIF and Pol II) transcription factors the DBPolF was formed on the AdMLP TATA box (Figure 6, lane 1). As expected, apart from TFIIA, all of these factors are indispensable for the formation of DBPolF complex since when one of them was omitted the complex did not form (Figure 6A, lanes 2–6; see also Cortes *et al.*, 1992; Flores *et al.*, 1992). An antibody raised against the CTD of the largest subunit of Pol II (mAb7G5), which inhibits transcription *in vitro* (data not shown), destroyed the DBPolF complex (Figure 6B, lane 6) further confirming the specificity of this complex. The Pol II fraction used in these experiments contains hTAF$_{II}$68 (Figure 5B, lane 4). Therefore, we tested whether hTAF$_{II}$68 was also incorporated in the DBPolF complex by adding the anti-TAF$_{II}$68 mAb (2B10)

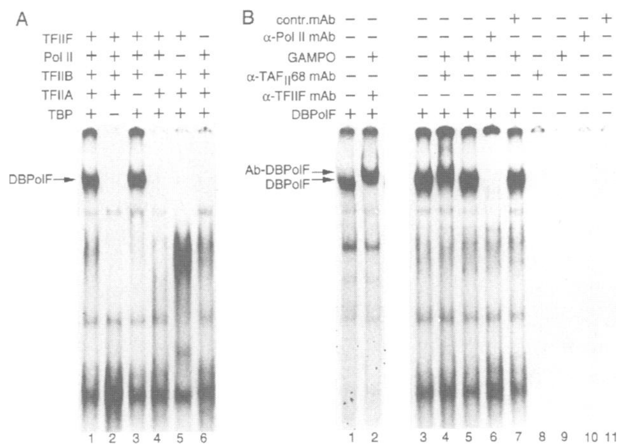


Fig. 6. hTAF$_{II}$68 incorporates into the PIC together with Pol II. (**A**) The protein–DNA complexes were assembled as described in Materials and methods and contained 50 ng of rTBP, rTFIIA subunits, rTFIIB, 3 μl of TFIIF and Pol II-containing fractions as indicated at the top of the panel. The position of the DNA–TBP–TFIIB–TFIIF–Pol II (DBPolF) complex is indicated. (**B**) The TATA box–DBPolF complex was assembled as in (A) and further incubated with the different antibodies as indicated at the top of the panel. The positions of the migration of the DBPolF complex and the corresponding complexes with either the anti-TAF$_{II}$68 mAb or with the anti-TFIIFβ mAb (Ab-DBPolF) are indicated. Note that to separate better the preinitiation complexes the free oligonucleotide probe was run out from the gel (see Materials and methods). GAMPO, peroxidase-conjugated goat anti-mouse IgG+IgM second antibodies .

to supershift the complex. The DBPolF complex has a molecular mass >750 kDa. To supershift this large protein–DNA complex by the anti-TAF$_{II}$68 mAb (2B10) it was first preincubated with peroxidase-conjugated goat anti-mouse antibodies (GAMPO). The anti-TAF$_{II}$68 antibody–GAMPO complex was able to supershift the DBPolF complex (Ab-DBPolF; Figure 6B, lane 4) while GAMPO alone had no effect (Figure 6B, lane 5). As a positive control a mAb directed against the small subunit of TFIIF (RAP30) was used. This anti-TFIIF mAb when preincubated with GAMPO supershifted the DBPolF complex to the same position as the anti-$TAF_{II}68$ antibody–GAMPO complex (compare lanes 2 and 4). Moreover, an antibody complex containing GAMPO and an unrelated mAb had no effect on the migration of the DBPolF complex (Figure 6B, lane 7). In the experiments described above rTBP was used and not TFIID. Therefore the only source of $hTAF_{II}68$ in these fractions is the Pol II fraction. The fact that the DBPolF preinitiation complex is supershifted by the anti-$TAF_{II}68$ mAb indicates that $hTAF_{II}68$ enters this complex together with Pol II and remains stably bound even under the stringent conditions of EMSA.

Discussion

Here we report the cloning and the characterization of a novel TFIID-associated factor, $hTAF_{II}68$, which has homology to human pro-oncoproteins and is associated with two multiprotein complexes involved in Pol II transcription, TFIID and the RNA polymerase II.

$hTAF_{II}68$ is a member of a growing sub-family of RNP-motif-containing proteins

RNP-motif-containing proteins have been described as being involved in numerous and diverse pathways of post-transcriptional regulation of gene expression (Burd *et al.*, 1994). We demonstrate for the first time that two different RNP-motif-containing proteins, $hTAF_{II}68$ and TLS/FUS, are present in Pol II transcription complexes indicating that these proteins may be involved in Pol II transcription. $hTAF_{II}68$ and TLS/FUS are members of the consensus RNA-binding domain (RNP-motif, RNP-CS) containing protein family. It is now clear that amongst the RNP-motif-containing proteins $hTAF_{II}68$ and TLS/FUS, together with EWS and the *Drosophila* Cabeza protein, belong to a new sub-family of proteins, which we have called the TET family (see Results and Figure 1C). All members of this growing sub-family of RNP-CS-containing proteins have an unusually large predicted loop structure (loop-2) between α-helix-1 and β-sheet-2 of the RNP-motifs and contain similar substitutions (Phe→Asp or Glu and Gly→Ala) in the highly conserved RNP-1 box (see Figure 1C and Crozat *et al.*, 1993).

TLS/FUS and EWS play a role in tumour formation following chromosomal translocations which result in the C-terminal half (including the RNA-binding domain) of the proteins being replaced by the DNA-binding domains of different transcription factors (Rabbitts, 1994). To understand how these chimeric fusion proteins induce tumour formation it is important to understand their normal function. The N-terminal ~200 amino acid regions of both TLS/FUS and EWS are glutamine-, serine- and tyrosine-rich and can act as strong transcriptional activators when fused to a DNA-binding domain (Ohno *et al.*, 1993, 1994; Prasad *et al.*, 1994; Sanchez-Garcia *et al.*, 1994). Similarly, the N-terminal region of $hTAF_{II}68$ contains a strong activation domain when fused to the DNA-binding domain of GAL4 (data not shown). Interestingly, some basal transcription factors have also been shown to activate transcription when fused to DNA-binding domains (Seipel *et al.*, 1993; Marsolier *et al.*, 1994; Chatterjee *et al.*, 1995; Klages *et al.*, 1995). These findings, together with the results presented in this study provide the first indication for the normal functions of the TET proteins in basal transcription.

The TFIID connection

Previously we have shown that functionally distinct TFIID complexes exist, composed of core and specific TAF_{II}s (see Introduction and Brou *et al.*, 1993a; Jacq *et al.*, 1994). To understand the functional differences observed for these TFIID complexes, TAF_{II}s that are unique for a distinct TFIID sub-population must be characterized. $hTAF_{II}68$ is a bona fide TAF_{II} based on the following lines of evidence: (i) it co-immunoprecipitates with TBP and $TAF_{II}100$ using highly stringent conditions (700 mM KCl washes); (ii) anti-$hTAF_{II}68$ mAbs co-imunoprecipitate both TBP and $TAF_{II}100$ and (iii) a highly related protein, TLS/FUS is also associated with a TFIID sub-population. The fact that anti-$hTAF_{II}68$ antibodies recognise $hTAF_{II}68$ mainly in the PC1.0-derived TFIID complex (Figure 3B) demonstrates that $hTAF_{II}68$ is a specific TAF_{II}. In addition, TLS/FUS is also a specific TAF_{II} since it was found to be associated with a functionally different TFIID sub-population, the PC0.3-derived TFIID complex (Figure 3B; and see also Brou *et al.*, 1993a; Jacq *et al.*, 1994).

Surprisingly, comparison of the amounts of core and specific TAF_{II}s associated with TBP versus the amounts of those TAF_{II}s present in TBP-free fractions revealed that amongst those tested only the core TAF_{II}, $hTAF_{II}100$, was almost exclusively associated with TBP (Figure 4, panels b and d). In contrast, a significantly large amount of core TAF_{II}s, such as $hTAF_{II}250$ and $hTAF_{II}55$, and also specific TAF_{II}s, such as $hTAF_{II}68$, TLS/FUS and $hTAF_{II}30$ were present in TBP-free fractions (Figure 4, panels a and c; and data not shown). These results suggest that most of the TAF_{II}s, including core and specific TAF_{II}s, may have several roles in the cell. In agreement with this, $hTAF_{II}250$ was first identified as a protein (CCG1) which plays a role in G_1 phase progression during the cell cycle (Sekiguchi *et al.*, 1991). However, it has not been determined whether the TBP-associated $hTAF_{II}250$ or $hTAF_{II}250$ present in the TBP-free fraction is involved in normal cell cycle progression. In the future it will be interesting to study the role of TAF_{II}s such as $hTAF_{II}68$, $hTAF_{II}55$ and $hTAF_{II}30$ in cellular processes other than transcription initiation.

Recently, putative substoichiometric *Drosophila* TAF_{II}s have been described that seem to play a role in determining promoter selectivity (Hansen *et al.*, 1995). The *Drosophila* protein, Cabeza (also called Sarcoma-Associated RNA-Binding Fly Homologue; Immanuel *et al.*, 1995; Stolow *et al.*, 1995) which also belongs to the same sub-family of RNP-motif-containing proteins as $hTAF_{II}68$, TLS/FUS and EWS (see Figure 1C), was shown to be associated with regions of chromatin transcribed by Pol II (Immanuel *et al.*, 1995). Thus, it would be of particular interest to test whether the 46 kDa substoichiometric $dTAF_{II}$ described by Hansen and Tjian (1995) would correspond to an isoform of Cabeza (SARFH type 2; Immanuel *et al.*, 1995) which has an apparent molecular weight of 45 kDa.

The presence of RNA and/or ssDNA-binding proteins in different TFIID complexes may provide new clues for the understanding of TFIID function. Similarly to

$hTAF_{II}68$, TLS/FUS and EWS have also been shown to bind ssDNA (Prasad *et al.*, 1994). Thus, these proteins may have a role during the PIC conversion from a 'closed' to an 'open' conformation. Moreover, $hTAF_{II}68$ and the other members of the TET sub-family may participate in defining the promoter selectivity of the distinct TFIID complexes. Alternatively it is also possible that these proteins by binding to the 5′ end of the newly synthesised RNA, play a role in RNA chain initiation and can make a bridge between the PIC and the elongation complex (see also below).

The RNA polymerase II connection

Several lines of evidence indicate that $hTAF_{II}68$ is also tightly associated with the Pol II: (i) a portion of $hTAF_{II}68$ co-purifies with Pol II on five subsequent chromatographic columns, (ii) $hTAF_{II}68$ cannot be dissociated from Pol II by high salt washes (1 M KCl; Figure 5C), the salt concentration that was shown to promote the dissociation of elongation factors, like TFIIF and TFIIS, from the elongation complexes (Flores *et al.*, 1989; Sopta *et al.*, 1989) and (iii) $hTAF_{II}68$ enters the PIC, together with Pol II and remains stably bound even under the stringent conditions of EMSA. The strong association of $hTAF_{II}68$ with the Pol II complex suggests that $hTAF_{II}68$ may be present in the elongation complex. Furthermore, $TAF_{II}68$ may also be a member of the human holo-polymerase II complex (Ossipow *et al.*, 1995). Anti-CTD antibodies are known to dissociate the core RNA polymerase II from the mediator complex constituting the yeast holo-polymerase complex (Kim *et al.*, 1994). In good agreement with the suggestion that $hTAF_{II}68$ is also a component of the human holo-polymerase complex the anti-CTD antibodies do not immunoprecipitate $hTAF_{II}68$ (data not shown) under the same conditions as those in which anti-$TAF_{II}68$ mAbs immunoprecipitate Pol II (Figure 5C).

To date two different TAF_{II}s, yTAF30 and $hTAF_{II}68$, have been identified as components of more than one transcription complex (Henry *et al.*, 1994; this study). These findings, together with the recent data that TAF_{II}s interact functionally with other basal transcription factors (Ruppert *et al.*, 1995; Dubrovskaya *et al.*, 1996) suggest that TAF_{II}s play an important role in the cross-talk between various components of the basal transcription machinery. Moreover, it is tempting to speculate that $hTAF_{II}68$ enters the PIC with TFIID and when elongation begins $hTAF_{II}68$ becomes associated with the elongating Pol II complex and/or with the nascent transcript.

Materials and methods

Purification of the TFIID complexes and polypeptide sequencing of $hTAF_{II}68$

Nuclear extracts from 1×10^{12} HeLa cells were prepared and purified on heparin Ultrogel, phosphocellulose P11 and phenyl-5PW columns as described in Brou *et al.* (1993a). For the preparative immunoprecipitation of TFIID complexes 80 ml (16 mg) of the PC1.0-derived phenyl 5PW fraction, containing TFIID, was used with 3 ml of protein G–Sepharose and 2 mg anti-TBP mAb, 3G3 (Brou *et al.*, 1993a). Bound TFIID was eluted by using an excess of the corresponding epitope peptide. The eluted TFIID was dialysed against buffer S (5 mM NaCl, 0.5 mM Tris–HCl pH 7.9, 5 μM DTT, 0.01% SDS) for 6 h, lyophilized, resolved by SDS–PAGE, transferred to a PVDF membrane and the 68 kDa polypeptide was digested with trypsin and microsequenced as described by (Brou *et al.*, 1993b).

Cloning of $hTAF_{II}68$ cDNA

The three peptide sequences underlined in Figure 1B were used to design degenerate oligonucleotide probes for screening a HeLa cell oligo(dT) primed cDNA library made in the λZAP II system according to Stratagene. After the third round of screening three positive clones were further analysed and sequenced. The longest of them had an insert of 2.1 kb, and contained an ORF encoding 589 amino acids. Three peptide sequences from tryptic digest of $hTAF_{II}68$ were found within this ORF.

Expression of recombinant $hTAF_{II}68$

The $hTAF_{II}68$ cDNA was excised from the BSK+ vector by *Pst*I–*Kpn*I digest and inserted in the same sites of the pAcSG-His NT-C vector (PharMingene). SF9 cell infection and whole cell extract preparation were performed as described in O'Reilly *et al.* (1992). The His-$TAF_{II}68$ was 90% pure in the SF9 cell pellet (Figure 2, lane 1). The cell pellet was dissolved in SDS sample buffer, boiled and loaded on SDS–PAGE. To construct the GST–RNP expression vector a partial $hTAF_{II}68$ cDNA clone (encoding for amino acids 175–414) was inserted in the *Sma*I–*Xho*I sites of the pGEX-4T3 vector (Pharmacia). GST protein overexpression and purification was as described in Smith and Johnson (1988). *In vitro* transcription–translation was performed by standard methods using the rabbit reticulocyte lysate kit (Promega).

Immunization and monoclonal antibody production

Immunization and monoclonal antibody production were essentially as described by Brou *et al.*, (1993a). Mice were injected intraperitoneally three times at 2 week intervals either with 100 μg of the synthetic peptides PC 84 ($hTAF_{II}68$ amino acids from 213 to 229) or PC183 (TLS/FUS amino acids from 266 to 283; see Crozat *et al.*, 1993) coupled to ovalbumin or purified GST–RNP (of $hTAF_{II}68$) fusion protein. Mice injected with $hTAF_{II}68$ synthetic peptide-coupled ovalbumin generated mAb2B11 and GST–RNP (of $hTAF_{II}68$) injected mice resulted in mAb2B10. Mice injected with TLS/FUS peptide-coupled ovalbumin generated mAb1D2. Spleen cells were fused with Sp2/0.Ag 14 myeloma cells and culture supernatants at day 10 were tested on recombinant and/or endogenous $hTAF_{II}68$ proteins by Western blot analysis, immunostaining of transfected Cos-1 cells and ELISA. The anti-TBP mAb (3G3), the anti-CTD mAb and the anti-$hTAF_{II}100$ mAbs (1TA and 2D2) were described previously (Brou *et al.*, 1993a; Besse *et al.*, 1995; Dubrovskaya *et al.*, 1996). The anti-hRPB 25 mAb was raised against the *E.coli* overexpressed and purified 25 kDa subunit of Pol II (a kind gift from J.Acker and M.Vigneron). The anti-TFIIFβ (RAP30) mAB was raised against the *E.coli* overexpressed small subunit of TFIIF (a kind gift from C.Chalut and J.M.Egly).

Immunoprecipitation and Western blot analysis

MAbs were purified from ascites fluids using caprylic acid precipitation followed by precipitation with 50% ammonium sulphate and dialysed. Immunoprecipitation was performed as described (Brou *et al.*, 1993a). Routinely 0.5–1 ml of the indicated fractions were immunoprecipitated with 50–100 μl protein G–Sepharose (Pharmacia) and 25 μg of the different mAbs as indicated in the legends to the figures. Antibody–protein G–Sepharose bound complexes were washed three times with IP buffer [25 mM Tris–HCl pH 7.9, 10% (v/v) glycerol, 0.1% NP40, 0.5 mM DTT, 5 mM $MgCl_2$] containing 700 mM (or more) KCl and twice with IP buffer containing 100 mM KCl. After washing bound proteins were either eluted with an excess of the corresponding epitope peptide and analysed by SDS–PAGE or 2–10 μl beads were boiled directly in SDS sample buffer and protein analysed on SDS–PAGE. Protein samples were then transferred to nitrocellulose membrane, and probed with the indicated antibody. The detection using an ECL kit (Amersham) was by standard methods.

South- and North-Western blot analysis

The RNA probe was synthesized with T7 RNA polymerase from the *Nco*I digested pAL4 plasmid (Ponglikitmongkol *et al.*, 1990) using [α-^{32}P]CTP and NTPs. The ss- and dsDNA probes were labelled using [γ-^{32}P]ATP and T4-polynucleotide kinase. Purified proteins were resolved on a 10% SDS–PAGE and either stained with Coomassie Blue (CBB) or transferred to a nitrocellulose membrane. The blots were prehybridized for 30 min at room temperature in buffer H (20 mM Tris–HCl pH 7.9, 50 mM NaCl, 1 mM DTT, 6 mM $MgCl_2$) containing 1% BSA, 20 μg/ml yeast transfer RNA and 3 U/ml RNasin. Hybridization was done for 1 h under the same conditions with 1.5×10^6 c.p.m./ml with each of the above described probes. The blots were washed three times for 15 min in buffer H and exposed to X-ray films for the appropriate time.

Purification of Pol II and basal transcription factors

Pol II was purified as follows (all procedures were performed at 4°C): the heparin-Ultrogel 0.6 M KCl-derived phosphocellulose 0.5 M KCl (PC0.5) fraction (Brou *et al.*, 1993a) was loaded on a DEAE-5PW HPLC column in 50 mM KCl containing buffer A [25 mM Tris–HCl pH 7.9, 0.5 mM DTT, 0.1 mM EDTA, 10% glycerol (v/v)] and eluted by steps with the same buffer containing 0.15 M, 0.25 M and 1 M KCl. Pol II was eluted between 0.25 and 1 M KCl and the Pol II-containing fractions were not contaminated by other basal transcription factors as verified by Western blot analysis (data not shown). To further purify Pol II, the DEAE-5PW HPLC fractions containing Pol II were pooled and loaded on a phenyl-5PW HPLC column in buffer A containing 0.9 M ammonium sulphate and proteins were eluted with a linear 0.9–0 M ammonium sulphate gradient. Pol II eluted in fractions containing ~0.12 M ammonium sulphate. The Pol II-containing fractions were reconcentrated on a DEAE-5PW HPLC column as described before.

The *E.coli* overexpression and the purification of recombinant human TBP, His-TFIIA α/β, His-TFIIAγ and TFIIB, were described previously (Moncollin *et al.*, 1992; Brou *et al.*, 1993a; Sun *et al.*, 1994). To purify TFIIF the PC0.5 fraction (Brou *et al.*, 1993a) was loaded on a phenyl-5PW HPLC column in buffer A containing 0.9 M ammonium sulphate and proteins were eluted with a linear 0.9–0 M ammonium sulphate gradient. TFIIF eluted in fractions containing ~0.7 M ammonium sulphate and were not contaminated by other basal transcription factors as verified by Western blot analysis (data not shown). All protein fractions were dialysed against buffer B [25 mM Tris–HCl pH 7.9, 50 mM KCl, 0.5 mM DTT, 0.1 mM EDTA, 20% glycerol (v/v)].

Electrophoretic mobility shift assay (EMSA)

For the ssDNA-binding specificity determination (Figure 2B), 20 ng of the GST–RNP protein was incubated for 15 min at 30°C with 50 fmol (20 000 c.p.m.) of a single stranded ^{32}P-5′-end labelled oligonucleotide, containing the AdMLP sequences from –40 to –11. Where indicated, the recombinant protein was preincubated with either mAb2B10 or with a 20, 500 or 2000 molar excess of deoxy-oligonucleotide homopolymers or with 2000 molar excess of ds oligo(dG/dC) (lane 19) or tRNA (lane 20). 20 µl reaction mixtures contained final concentrations of 12 mM HEPES pH 8, 60 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 0.5 mM EDTA, 0.05% NP40, 10% glycerol (v/v) and 1 µg BSA. The complexes were separated by electrophoresis through a 5% polyacrylamide gel using 0.5×TBE buffer [50 mM Tris base, 50 mM boric acid and 1 mM EDTA (pH 8.3)] at 120 V. The gels were then dried and subjected to autoradiography.

In Figure 6, protein components (as indicated) were first preincubated for 10 min on ice with 500 ng poly(dG-dC)/(dG-dC). Then 50 fmol (20 000 c.p.m.) of a double stranded ^{32}P-5′-end labelled oligonucleotide, containing the AdMLP sequences from –40 to –11, was added to the reactions and incubated for 30 min at 30°C. Where indicated (see Figure 6B) antibodies were added and the reactions were further incubated for another 15 min at 30°C. Electrophoresis was carried out as above except that the free probe was run out of the gel.

Acknowledgements

We are grateful to C.Brou for the initial large scale TFIID purification, to T.Sekiguchi and T.Nishimoto for the anti-CCG1 serum, to J.-M.Garnier for the gift of cDNA libraries, to J.Acker and M.Vigneron for Pol II antibodies, to A.C.Lavigne, R.Roy and Y.Trottier for helpful discussions and to R.A.Fraser and I.Davidson for critical reading of the manuscript. We thank E.Scheer for excellent technical assistance, A.Staub for peptide microsequencing, P.Eberling for peptide synthesis, S.Vicaire for sequencing, I.Cheynel for expression of His-TAF$_{II}$68 in the baculovirus system, N.Jung and V.Schultz for technical help in the antibody preparations, the cell culture group for HeLa cells, C.Werlé, J.-M.Lafontaine and B.Boulay for illustrations and photography. D.J.H. is supported by an European Molecular Biology Organization Long Term Postdoctoral Fellowship. The present work was supported by grants from the CNRS, the INSERM, Centre Hospitalier Universitaire Regional, the Ministère de la Recherche et Technologie, Fondation pour la Recherche Médicale and the Association pour la Recherche contre le Cancer.

References

Besse,S., Vigneron,M., Pichard,E. and Puvion-Dutilleul,F. (1995) Synthesis and maturation of viral transcripts in herpes simplex virus type 1 infected HeLa cells: the role of interchromatin granules. *Gene Expr.*, **4**, 143–161.

Brou,C., Chaudhary,S., Davidson,I., Lutz,Y., Wu,J., Egly,J.M., Tora,L. and Chambon,P. (1993a) Distinct TFIID complexes mediate the effect of different transcriptional activators. *EMBO J.*, **12**, 489–499.

Brou,C., Kuhn,A., Staub,A., Chaudhary,S., Grummt,I., Davidson,I. and Tora,L. (1993b) Sequence-specific transactivators counteract topoisomerase II-mediated inhibition of *in vitro* transcription by RNA polymerase I and II. *Nucleic Acids Res.*, **21**, 4011–4018.

Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

Chatterjee,S. and Struhl,K. (1995) Connecting a promoter-bound protein to TBP bypasses the need for a transcriptional activation domain. *Nature*, **374**, 820–822.

Chen,J.L., Attardi,L.D., Verrijzer,C.P., Yokomori,K. and Tjian,R. (1994) Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. *Cell*, **79**, 93–105.

Chiang,C.M. and Roeder,R.G. (1995) Cloning of an intrinsic human TFIID subunit that interacts with multiple transcriptional activators. *Science*, **267**, 531–536.

Chiang,C.M., Ge,H., Wang,Z., Hoffmann,A. and Roeder,R.G. (1993) Unique TATA-binding protein-containing complexes and cofactors involved in transcription by RNA polymerases II and III. *EMBO J.*, **12**, 2749–2762.

Cortes,P., Flores,O. and Reinberg,D. (1992) Factors involved in specific transcription by mammalian RNA polymerase II: purification and analysis of transcription factor IIA and identification of transcription factor IIJ. *Mol. Cell. Biol.*, **12**, 413–421.

Crozat,A., Aman,P., Mandahl,N. and Ron,D. (1993) Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. *Nature*, **363**, 640–644.

Delattre,O. *et al.* (1992) Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature*, **359**, 162–165.

Dubrovskaya,V., Lavigne,A.-C., Davidson,I., Acker,J., Staub,A. and Tora,L. (1996) Distinct domains of hTAF$_{II}$100 are required for functional interaction with transcription factor TFIIFβ (RAP30) and incorporation into the TFIID complex. *EMBO J.*, **15**, 3702–3712.

Flores,O., Maldonado,E. and Reinberg,D. (1989) Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J. Biol. Chem.*, **264**, 8913–8921.

Flores,O., Lu,H. and Reinberg,D. (1992) Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *J. Biol. Chem.*, **267**, 2786–2793.

Goodrich,J.A., Hoey,T., Thut,C.J., Admon,A. and Tjian,R. (1993) *Drosophila* TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. *Cell*, **75**, 5–30.

Hansen,S.K. and Tjian,R. (1995) TAFs and TFIIA mediate differential utilization of the tandem Adh promoters. *Cell*, **82**, 565–575.

Hengartner,C.J., Thompson,C.M., Zhang,J., Chao,D.M., Liao,S.M., Koleske,A.J., Okamura,S. and Young,R.A. (1995) Association of an activator with an RNA polymerase II holoenzyme. *Genes Dev.*, **9**, 897–910.

Henry,N.L., Campbell,A.M., Feaver,W.J., Poon,D., Weil,P.A. and Kornberg,R.D. (1994) TFIIF-TAF-RNA polymerase II connection. *Genes Dev.*, **8**, 2868–2878.

Histake,K., Ohta,T., Takada,R., Guermah,M., Horikoshi,M., Nakatani,Y. and Roeder,R.G. (1995) Evolutionary conservation of human TATA-binding polypeptide-associated factors TAFII31 and TAFII80 and interactions of TAFII80 with other TAFs and with general transcription factors. *Proc. Natl Acad. Sci. USA*, **92**, 8195–8199.

Hoey,T., Weinzierl,R.O., Gill,G., Chen,J.L., Dynlacht,B.D. and Tjian,R. (1993) Molecular cloning and functional analysis of *Drosophila* TAF110 reveal properties expected of coactivators. *Cell*, **72**, 247–260.

Holstege,F.C., Tantin,D., Carey,M., van der Vliet,P.C. and Timmers,H.T. (1995) The requirement for the basal transcription factor IIE is determined by the helical stability of promoter DNA. *EMBO J.*, **14**, 810–819.

Immanuel,D., Zinszner,H. and Ron,D. (1995) Association of SARFH (sarcoma-associated RNA-binding fly homolog) with regions of chromatin transcribed by RNA polymerase II. *Mol. Cell. Biol.*, **15**, 4562–4571.

Jacq,X., Brou,C., Lutz,Y., Davidson,I., Chambon,P. and Tora,L. (1994) Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell*, **79**, 107–117.

Kane,C.M. (1994) Transcription elongation and gene regulation in eukaryotes. In Conaway,R.C. and Conaway,J.W. (eds), *Transcription Mechanisms and Regulation*. Raven Press, New York, pp. 279–296.

Kenan,D.J., Query,C.C. and Keene,J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem. Sci.*, **16**, 214–220.

Kiledjian,M. and Dreyfuss,G. (1992) Primary structure and binding activity of the hnRNP U protein: binding RNA through RGG box. *EMBO J.*, **11**, 2655–2664.

Kim,Y.J., Bjorklund,S., Li,Y., Sayre,M.H. and Kornberg,R.D. (1994) A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, **77**, 599–608.

Klages,N. and Strubin,M. (1995) Stimulation of RNA polymerase II transcription initiation by recruitment of TBP *in vivo*. *Nature*, **374**, 822–823.

Klemm,R.D., Goodrich,J.A., Zhou,S. and Tjian,R. (1995) Molecular cloning and expression of the 32-kDa subunit of human TFIID reveals interactions with VP16 and TFIIB that mediate transcriptional activation. *Proc. Natl Acad. Sci. USA*, **92**, 5788–5792.

Marsolier,M.C., Chaussivert,N., Lefebvre,O., Conesa,C., Werner,M. and Sentenac,A. (1994) Directing transcription of an RNA polymerase III gene via GAL4 sites. *Proc. Natl Acad. Sci. USA*, **91**, 138–142.

Mengus,G., May,M., Jacq,X., Staub,A., Tora,L., Chambon,P. and Davidson,I. (1995) Cloning and characterization of hTAFII18, hTAFII20 and hTAFII28: three subunits of the human transcription factor TFIID. *EMBO J.*, **14**, 1520–1531.

Moncollin,V., Schaeffer,L., Chalut,C. and Egly,J.M. (1992) Expression in *Escherichia coli*: purification and properties of the recombinant human general transcription factor rTFIIB. *Protein Expr. Purif.*, **3**, 374–379.

Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.

O'Reilly,D.R., Miller,L.K. and Luckow,V.A. (1992) Baculovirus Expression Vectors. Freedman,W.H., New York, pp. 1–341.

Ohno,T., Rao,V.N. and Reddy,E.S. (1993) EWS/Fli-1 chimeric protein is a transcriptional activator. *Cancer Res.*, **53**, 5859–5863.

Ohno,T., Ouchida,M., Lee,L., Gatalica,Z., Rao,V.N. and Reddy,E.S. (1994) The EWS gene, involved in Ewing family of tumors, malignant melanoma of soft parts and desmoplastic small round cell tumors, codes for an RNA-binding protein with novel regulatory domains. *Oncogene*, **9**, 3087–3097.

Ossipow,V., Tassan,J.P., Nigg,E.A. and Schibler,U. (1995) A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell*, **83**, 137–146.

Parvin,J.D., Shykind,B.M., Meyers,R.E., Kim,J. and Sharp,P.A. (1994) Multiple sets of basal factors initiate transcription by RNA polymerase II. *J. Biol. Chem.*, **269**, 18414–18421.

Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.

Ponglikitmongkol,M., White,J.H. and Chambon,P. (1990) Synergistic activation of transcription by the human estrogen receptor bound to tandem responsive elements. *EMBO J.*, **9**, 2221–2231.

Prasad,D.D., Ouchida,M., Lee,L., Rao,V.N. and Reddy,E.S. (1994) TLS/FUS fusion domain of TLS/FUS-erg chimeric protein resulting from the t(16; 21) chromosomal translocation in human myeloid leukemia functions as a transcriptional activation domain. *Oncogene*, **9**, 3717–3729.

Pugh,B.F. and Tjian,R. (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.*, **5**, 1935–1945.

Rabbitts,T.H. (1994) Chromosomal translocations in human cancer. *Nature*, **372**, 143–149.

Ruppert,S., Wang,E.H. and Tjian,R. (1993) Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation. *Nature*, **362**, 175–179.

Ruppert,S. and Tjian,R. (1995) Human TAFII250 interacts with RAP74: implication for RNA polymerase II initiation. *Genes Dev.*, **9**, 2747–2755.

Sanchez-Garcia,I. and Rabbitts,T.H. (1994) Transcriptional activation by TAL1 and FUS-CHOP proteins expressed in acute malignancies as a result of chromosomal abnormalities. *Proc. Natl Acad. Sci. USA*, **91**, 7869–7873.

Sauer,F., Hansen,S.K. and Tjian,R. (1995) Multiple TAFIIs directing synergistic activation of transcription. *Science*, **270**, 1783–1788.

Seipel,K., Georgiev,O., Gerber,H.P. and Schaffner,W. (1993) C-terminal domain (CTD) of RNA-polymerase II and N-terminal segment of the human TATA-binding protein (TBP) can mediate remote and proximal transcriptional activation, respectively. *Nucleic Acids Res.*, **21**, 5609–5615.

Sekiguchi,T., Nohiro,Y., Nakamura,Y., Hisamoto,N. and Nishimoto,T. (1991) The human CCG1 gene, essential for progression of the G1 phase, encodes a 210-kilodalton nuclear DNA-binding protein. *Mol. Cell. Biol.*, **11**, 3317–3325.

Serizawa,H., Conaway,J.W. and Conaway,R.C. (1994) Transcriptional initiation by mammalian RNA polymerase II. In Conaway,R.C. and Conaway,J.W. (eds), *Transcription Mechanisms and Regulation*. Raven Press, New York, pp. 27–43.

Smith,D.B. and Johnson,K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. *Gene*, **67**, 31–40.

Sopta,M., Burton,Z.F. and Greenblatt,J. (1989) Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II. *Nature*, **341**, 410–414.

Stolow,D.T. and Haynes,S.R. (1995) Cabeza, a *Drosophila* gene encoding a novel RNA-binding protein, shares homology with EWS and TLS, two genes involved in human sarcoma formation. *Nucleic Acids Res.*, **23**, 835–843.

Sun,X., Ma,D., Sheldon,M., Yeung,K. and Reinberg,D. (1994) Reconstitution of human TFIIA activity from recombinant polypeptides: a role in TFIID-mediated transcription. *Genes Dev.*, **8**, 2336–2348.

Tanese,N., Pugh,B.F. and Tjian,R. (1991) Coactivators for a proline-rich activator purified from the multisubunit human TFIID complex. *Genes Dev.*, **5**, 2212–2224.

Timmers,H.T. (1994) Transcription initiation by RNA polymerase II does not require hydrolysis of the beta-gamma phosphoanhydride bond of ATP. *EMBO J.*, **13**, 391–399.

Timmers,H.T., Meyers,R.E. and Sharp,P.A. (1992) Composition of transcription factor B-TFIID. *Proc. Natl Acad. Sci. USA*, **89**, 8140–8144.

Verrijzer,C.P., Yokomori,K., Chen,J.L. and Tjian,R. (1994) *Drosophila* TAFII150: similarity to yeast gene TSM-1 and specific binding to core promoter DNA. *Science*, **264**, 933–941.

Weinzierl,R.O., Ruppert,S., Dynlacht,B.D., Tanese,N. and Tjian,R. (1993) Cloning and expression of Drosophila TAFII60 and human TAFII70 reveal conserved interactions with other subunits of TFIID. *EMBO J.*, **12**, 5303–5309.

White,J., Brou,C., Wu,J., Lutz,Y., Moncollin,V. and Chambon,P. (1992) The acidic transcriptional activator GAL-VP16 acts on preformed template-committed complexes. *EMBO J.*, **11**, 2229–2240.

Yankulov,K., Blau,J., Purton,T., Roberts,S. and Bentley,D.L. (1994) Transcriptional elongation by RNA polymerase II is stimulated by transactivators. *Cell*, **77**, 749–759.

Yokomori,K., Chen,J.-L., Admon,A., Zhou,S. and Tjian,R. (1993) Molecular cloning and characterization of dTAF$_{II}$30a and dTAF$_{II}$30b: two small subunits of Drosophila TFIID. *Genes Dev.*, **7**, 2587–2597.

Zawel,L. and Reinberg,D. (1995) Common themes in assembly and function of eukaryotic transcription complexes. *Annu. Rev. Biochem.*, **64**, 533–561.

Zhou,Q., Lieberman,P.M., Boyer,T.G. and Berk A.J. (1992) Holo-TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. *Genes Dev.*, **6**, 64–74.

Received on March 20, 1996; revised on June 11, 1996

Note added in proof

The accession number in the EMBL Nucleotide Sequence Database for the sequence reported in this paper is X98893.